HLT-NAACL 06

TextGraphs: Graph-based algorithms for Natural Language Processing

Proceedings of the Workshop

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53704

©2006 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

PREFACE

Graph theory is a well studied discipline, and so is the field of natural language processing. Traditionally, these two areas of study have been perceived as distinct, with different algorithms, different applications, and different potential end-users. However, as recent research work has shown, the two disciplines are in fact intimately connected, with a large variety of natural language processing applications finding efficient solutions within graph-theoretical frameworks.

This volume contains papers accepted for presentation at the Textgraphs 2006 Workshop on Graph-based Algorithms for Natural Language Processing. This event took place on June 9, 2006, in New York City, immediately following the HLT-NAACL Human Language Technologies Conference. The workshop was centered around the topic of using graph-based algorithms for natural language processing, and it brought together people working on areas as diverse as lexical semantics, text summarization, text mining, ontology construction, clustering and learning, connected by the common underlying theme consisting of the use of graph-theoretical methods for text processing tasks.

We issued calls for both regular and short, late–breaking papers. After careful review by our program committee, eleven regular papers and four short papers were accepted for presentation. We were truly impressed by the high quality of the reviews provided by all the members of the program committee, particularly since deadlines were very tight. All of the committee members provided timely and thoughtful reviews, and the papers that appear have certainly benefited from that expert feedback.

Finally, when we first started planning this workshop, we agreed that having a high quality invited speaker was crucial. We thank Lillian Lee not only for her talk, but also for the boost of confidence provided by her quick and enthusiastic acceptance.

Rada Mihalcea and Dragomir Radev
June 2006

CHAIRS:

Dragomir Radev, University of Michigan
Rada Mihalcea, University of North Texas

INVITED SPEAKER:

Lillian Lee, Cornell University

PROGRAM COMMITTEE:

Lada Adamic, University of Michigan
Răzvan Bunescu, University of Texas at Austin
Timothy Chklovski, USC / Information Sciences Institute
Diane Cook, University of Texas at Arlington
Inderjit Dhillon, University of Texas at Austin
Beate Dorow, University of Stuttgart
Gael Dias, Universidade da Beira Interior Portugal
Kevin Gee, University of Texas at Arlington
Lise Getoor, University of Maryland
Güneş Erkan, University of Michigan
John Lafferty, Carnegie Mellon University
Lillian Lee, Cornell University
Andrew McCallum, University of Massachusetts
Bo Pang, Cornell University
Patrick Pantel, USC / Information Sciences Institute
Paul Tarau, University of North Texas
Simone Teufel, University of Cambridge
Lucy Vanderwende, Microsoft Research
Florian Wolf, F-W Consulting
Dominic Widdows, Maya Design
Hongyuan Zha, Penn State
Xiaojin Zhu, University of Wisconsin

WEBSITE:

http://www.textgraphs.org/ws06

Table of Contents

A Graphical Framework for Contextual Search and Name Disambiguation in Email
Einat Minkov, William Cohen and Andrew Ng 1

Graph Based Semi-Supervised Approach for Information Extraction
Hany Hassan, Ahmed Hassan and Sara Noeman 9

Graph-Based Text Representation for Novelty Detection
Michael Gamon 17

Measuring Aboutness of an Entity in a Text
Marie-Francine Moens, Patrick Jeuniaux, Roxana Angheluta and Rudradeb Mitra 25

A Study of Two Graph Algorithms in Topic-driven Summarization
Vivi Nastase and Stan Szpakowicz 29

Similarity between Pairs of Co-indexed Trees for Textual Entailment Recognition
Fabio Massimo Zanzotto and Alessandro Moschitti 33

Learning of Graph-based Question Answering Rules
Diego Molla 37

Seeing stars when there arent many stars: Graph-based semi-supervised learning for sentiment categorization
Andrew Goldberg and Xiaojin Zhu 45

Random-Walk Term Weighting for Improved Text Classification
Samer Hassan and Carmen Banea 53

Graph-based Generalized Latent Semantic Analysis for Document Representation
Irina Matveeva and Gina-Anne Levow 61

Synonym Extraction Using a Semantic Distance on a Dictionary
Philippe Muller, Nabil Hathout and Bruno Gaume 65

Chinese Whispers - an Efficient Graph Clustering Algorithm and its Application to Natural Language Processing Problems
Chris Biemann 73

Matching syntactic-semantic graphs for semantic relation assignment
Vivi Nastase and Stan Szpakowicz 81

Evaluating and optimizing the parameters of an unsupervised graph-based WSD algorithm
Eneko Agirre, David Martínez, Oier López de Lacalle and Aitor Soroa 89

Context Comparison as a Minimum Cost Flow Problem
Vivian Tsang and Suzanne Stevenson 97

Conference Program

Friday, June 9, 2006

08:45–09:00 Introduction

Session 1: Session One

09:00–09:25 *A Graphical Framework for Contextual Search and Name Disambiguation in Email*
Einat Minkov, William Cohen and Andrew Ng

09:25–09:50 *Graph Based Semi-Supervised Approach for Information Extraction*
Hany Hassan, Ahmed Hassan and Sara Noeman

09:50–10:15 *Graph-Based Text Representation for Novelty Detection*
Michael Gamon

10:15–10:30 *Measuring Aboutness of an Entity in a Text*
Marie-Francine Moens, Patrick Jeuniaux, Roxana Angheluta and Rudradeb Mitra

10:30–11:00 Coffee break

Session 2: Session Two

11:00–11:15 *A Study of Two Graph Algorithms in Topic-driven Summarization*
Vivi Nastase and Stan Szpakowicz

11:15–11:30 *Similarity between Pairs of Co-indexed Trees for Textual Entailment Recognition*
Fabio Massimo Zanzotto and Alessandro Moschitti

11:30–12:30 Invited talk ”Sense and Sensibility” by Lillian Lee

12:30–14:00 Lunch break

Friday, June 9, 2006 (continued)

Session 3: Session Three

14:00–14:25 *Learning of Graph-based Question Answering Rules*
Diego Molla

14:25–14:50 *Seeing stars when there arent many stars: Graph-based semi-supervised learning for sentiment categorization*
Andrew Goldberg and Xiaojin Zhu

14:50–15:15 *Random-Walk Term Weighting for Improved Text Classification*
Samer Hassan and Carmen Banea

15:15–15:30 *Graph-based Generalized Latent Semantic Analysis for Document Representation*
Irina Matveeva and Gina-Anne Levow

15:30–16:00 Coffee break

Session 4: Session Four

16:00–16:25 *Synonym Extraction Using a Semantic Distance on a Dictionary*
Philippe Muller, Nabil Hathout and Bruno Gaume

16:25–16:50 *Chinese Whispers - an Efficient Graph Clustering Algorithm and its Application to Natural Language Processing Problems*
Chris Biemann

16:50–17:15 *Matching syntactic-semantic graphs for semantic relation assignment*
Vivi Nastase and Stan Szpakowicz

17:15–17:40 *Evaluating and optimizing the parameters of an unsupervised graph-based WSD algorithm*
Eneko Agirre, David Martínez, Oier López de Lacalle and Aitor Soroa

17:40–18:05 *Context Comparison as a Minimum Cost Flow Problem*
Vivian Tsang and Suzanne Stevenson